心理科学进展 2023, Vol. 31, No. 6, 1–9
Advances in Psychological Science

https://doi.org/10.3724/SP.J.1042.2023.001

# 预测加工与认知边界：基于现象学的视角*

李雪玉 高申春

(吉林大学哲学社会学院，长春 130012)

**摘 要** 预测加工是目前认知科学和心灵哲学中一个重要的理论框架。该框架对于认知边界有两种截然不同的理解方式，分为内在主义和外在主义立场。前者主张以马尔科夫毯为证据边界，认知是内在的心智活动，是一种彻底的神经表征主义；而后者倡导该框架与具身心智、延展认知的观点相结合，以行动为导向的表征使心物之间并无静态固定的边界。两者均面临无法解释的疑难，认知限于颅内则遭遇认识的超越性困境，而动态可变的边界则受困于认知标志的模糊性。现象学在认识的本质问题上作出了富有意义的探索，它无疑为预测加工认知模型提供了反思空间。

**关键词** 预测加工，认知边界，内在主义，外在主义，现象学

**分类号** B84-06



## 1 引言

预测加工(predictive processing, PP)是目前认知科学和心灵哲学中一个重要的理论框架(Hohwy, 2020; Piekarski, 2021)，它意味着一套研究人类主观经验的独特工具和概念，大脑被视为预测机器(prediction machine) (Clark, 2013; Prinz, 2019; Seth, 2021)，能够利用已有的先天知识来预测即将到来的感官输入信息，形成一种关于外部世界状态的最佳假设。围绕该框架的探讨，尤其是关于主体与外部世界关系所开展的研究，成为认知的内在主义与外部主义争论的新战场。我们作为认知的主体是如何认识外部世界的？是否存在“认知边界” (Adams & Aizawa, 2008; Clark & Chalmers, 1998; Kersten, 2022; Ramstead et al., 2021)？PP框架对此有两种截然不同的理解方式。一种支持认识的内在主义立场，将预测机制当作认知者内在的心智活动，从而将认知限于以神经为中心的边界之内，并且认为人们所直接把握的是“脑中的世界模型” (Frith, 2007)或某种“虚拟现实” (Hohwy, 2013)。而另一种解释则以行动为导向，主张认知并非限于颅内，它依赖于主体之外的关系或性质，该框架使得我们更为直接地接触到外部“世界的本来面目” (安迪·克拉克, 2020)。对以上两种解释进行细致分析发现，他们的争论正值白热化时期，断然在两者之间作出非此即彼的选择几乎是不可能的，同时在说明认知的本质时他们均面临着不可摆脱的冲突。伴随着认知科学对意识的兴趣复苏、大脑神经科学以及具身认知的发展，我们无法再对认识的内在和超越问题漠不关心，而现象学在这些问题上作出了富有意义的探索。或许正因如此，现象学的思想被应用于意识、知觉和身体的认知研究之中，2002年《现象学与认知科学》(*Phenomenology and the Cognitive Sciences*)创刊，成为这两个领域合作研究的重要学术阵地。在新近的研究中，Gallagher和Zahavi (2021)愈加相信“现象学哲学可以解决当代心灵哲学和认知科学中的核心问题”，并且有研究者(Nave, 2022)采用Husserl和Merleau-Ponty等现象学家的见解来思考PP模型的研究。鉴于此，本研究尝试基于现象学的视角，分析PP认知模型及其内在主义与外在主义解释所遭致的疑难，力图修补和完善静默于其中的根源问题的基本主张。

## 2 PP认知模型与马尔可夫毯

### 2.1 PP作为一种认知模型

预测加工也常被称为预测大脑(predictive

收稿日期：2022-09-08
通信作者：高申春, E-mail: gsc@jlu.edu.cn

brain)、预测心智(predictive mind)、预测编码(predictive coding)或预测误差最小化(prediction error minimization)。它是“关于心智的新科学” (安迪·克拉克, 2020), 是人们探索新的认知观的一种表达。有研究者认为, 该框架的感知观实现了一种“哥白尼反转” (Copernican inversion) (Seth, 2021)。在传统的认知科学观点中, 大脑是刺激驱动的被动装置, 感知内容从传入的的感官信号中以自下而上或由外而内的方式呈现, 这一观点(至少在表面上)是与(计算机)实验相符的。而在PP认知模型中, 我们在外界刺激传入之前已进行某种预测或猜想, 感知内容是由大脑对感觉输入的隐藏原因的最佳猜测构成的, 这种最佳猜测也被 Dołega 和 Dewhurst (2021)称作获胜假设, 即具有最高后验概率的假设。在感知层级结构中自上而下的信号传递预测, 自下而上的信号传递预测误差, 认知便成为一种不断将预测误差最小化的过程。

不可否认, 这种预测的感知观在一定程度上呼应了康德的某些洞见, Hohwy (2013)指出它具有一种“独特的康德元素”。的确, 我们不难发现, PP框架的核心观点展现了《纯粹理性批判》的思想, 与传统的说法迥然相异的是, 预测大脑运用自上而下的方式对认知进行逆向分析, 正如康德在认识论中实现的“哥白尼革命”, 我们不妨“假设对象必须符合我们的认识”, 这种认识“在对象被给予我们之前就对它们有所断定” (Kant, 1996)。需要说明的是, 我们有史以来的一个几乎可以说是坚不可摧的传统观念是, 首先确认外部世界具有预先给定的属性, 而认知是对这个外部世界的表征, 这表现为认识论中的符合论, 这种二元论的表征式认识论也被Dreyfus和Taylor (2015)理解为中介认识论(Mediational Epistemology)。Swanson (2016)认为该框架“以康德无法想象的方式超越了康德的见解”, 并非说 PP 对上述成见有所突破, 而是指它的研究方法(如, 概率、计算以及神经科学基础等方面)。Helmholtz 作为新康德主义者, 在19世纪生理学和心理学创立之时试图将康德的认识论问题向前推进一步, 因此有研究者认为PP最相关的历史可能出现在 Helmholtz 的“无意识推理” (unconscious inference) (Hohwy, 2013; Piekarski, 2021; Seth, 2021)的概念中, 他最先明确指出大脑是假说检验机器。他们的观点不谋而合, 感知是由内而外的一个积极建设性的过程, 而非对外部世界的被动呈现。以一朵红花的知觉为例子, 尽管信息由感官输入大脑的过程是自下而上的, 但我们需要依照某种规则对杂多表象进行选择和组合, 从而产生整朵红花的知觉, 自上而下的加工在此过程之中引导了直观对象的构造活动。

不仅如此, PP 的实证研究也试图证实以上的观点。在该模型研究的早期, Rao 和 Ballard (1999)的实验结果表明, 视觉中的超感受野效应并非仅仅是前馈现象, 还可能是由于视觉系统使用有效的分层策略对自然图像进行编码而产生的反馈, 在此视觉模型中, 神经活动从高阶到低阶视觉皮层的反馈携带着预测, 而前馈携带预测和实际低阶活动之间的误差。近来, 研究者们提出失匹配负波(mismatch negativity, MMN)是 PP 的预测误差的外部表现, 它是大脑的一种典型的电生理反应。Wacongne 等(2012)提出了基于预测编码的听觉皮层神经元模型, 可以解释对罕见异常的频率依赖性反应、对交替序列中意外重复的反应(ABABAA...)、缺乏对全局序列背景的考虑、对声音缺失的反应等 MMN 的主要经验特性。预测的观点不仅应用于外部感知(如视觉、听觉), 而且也被用来描述内感知, Seth (2013)指出“内感知推理”将主观感觉状态(情绪)视为源自对内感知传入原因的积极推断的生成(预测)模型, 认为情绪是从生理变化的认知评估中产生的, 它为身体所有权经验(拥有并认同某个特定的身体)和自我意识经验的神经认知机制提供了新的线索。安迪·克拉克(2020)指出“fMRI 和 EEG 领域越来越多的成果都支持 PP 模型所设想的各种关系”, 不过有研究者认为“关于 PP 的经验基础的争论目前尚未解决” (Walsh et al., 2020)。值得思考的是, 理论与实证的发展的同时也存在着紧张关系, 争论源自于几个有趣的问题和概念, 其中涉及一个核心的问题: 何为认知的本质? 认知的边界在哪里? 边界的位置决定了心智与世界的关系, 其中隐含着关于理解人类心智的秘密。

### 2.2 以马尔可夫毯为界?

如何理解PP认知边界问题, 将基于一些关于马尔可夫毯(Markov blanket)概念的假设。这一概念最初源于机器学习的文献中, Pearl (1988)将马尔可夫毯用于在统计上把贝叶斯网络中的一个节点与网络中的其它节点分离, 它由节点的父节点、子节点和子节点的其它父节点组成。随之, 它

被纳入预测加工的框架下并被用来探讨认知的预测机制是否应当被“限制于大脑颅腔之中” (Clark, 2017; Hohwy, 2016; Ramstead et al., 2021)。通过考察发现，该毯起初只是确定一个节点及其子节点与其它节点之间的区别以促进更精确的推理，是一种启发式方法；而在PP模型中它被作为一个本体论的概念，Friston (2010)采用“自由能原理”对生物系统存在的意义进行考察，认为预测框架下的认知是生命系统将理论上的意外事件最小化的结果，马尔科夫毯可以 “区分内部状态和外部状态”，同时“内部状态(及其毯子)似乎会参与主动贝叶斯推理” (Friston, 2013)。随之，Hohwy (2016)将马尔可夫毯的某些部分定位到了“脊髓背角”。Bruineberg 等(2021)将该毯称作“皇帝的新马尔可夫毯”，并建议作出其原始认知用法与新的形而上学结构的区分，将其最初概念与PP中的应用区分为“珍珠毯” (Pearl blankets)与“弗利斯顿毯” (Friston blankets)。那么，一个不容忽视的问题是，对原有概念扩展应用的依据何在？

据目前的研究而言，两种关于PP的解释有着相同的前提假设，即贝叶斯大脑猜测被马尔可夫毯遮挡住的隐藏环境变量。一方面，在某种意义上，马尔可夫毯的存在屏蔽了内部状态与外部状态的直接影响，外部通过毯子的状态间接地影响内部状态。有研究者提出，该毯由活动状态和感觉状态这两个部分组成，由此使得内部(系统)和环境(神经元、身体和世界)之间的互动成为可能。内部可以通过活动状态来影响外部，外部可以通过感觉状态影响内部，由此该毯“提供了描述生物和/或认知系统边界的统计手段” (Ramstead et al., 2021)。那么对于认为认知边界受大脑约束的解释来说，这便是一种标准的内在主义的观点，即 PP意味着“大脑本质上是自我证明的”，它与环境和身体其它部位的感官输入的外部原因之间形成“证据边界” (Hohwy, 2016)。与此不同的是，对于以行动为导向的解释来说，他们坚持认知的外部主义，该解释方案通过强调主动推理以整合大脑、身体和世界的紧密关系，换言之，马尔可夫毯并非心智和世界的固定边界，Fabry (2017)和 Clark (2017)持有类似的立场。

事实上，Friston 对马尔可夫毯的概念引入依据未置一词，然而我们不能将他的沉默视为一种默认。由此观之，与其说马尔科夫毯解释了认知边界的问题，不如说是转移了人们对真正问题的讨论而陷入了内部主义与外部主义的哲学辩论中。用 Metzinger (2017)的话来说，我们既要避免“穿着马尔可夫毯的俄罗斯套娃认识论”，也要防止为“延展心智争论”搭建舞台。

## 3 PP 模型内在主义解释及其困境

### 3.1 PP 模型的认知内在主义解释

恰如“内在”概念本身所意指的那样，认知的内在主义赋予主体内在状态以优先地位，我们所直接把握的是个人心中的内在状态。与之相似，PP模型“彻底的神经表征主义” (radical neuro-representationalism) (Gallagher & Zahavi, 2021)的观点无不体现着内在论的核心内容，我们通常将这类理解称为认知内在主义。那么，我们所知觉到的、我们所直接经验或接触到的究竟是什么呢？Frith (2007)在他关于预测大脑的著作中写道，“大脑的世界模型才是我们真正知觉到的东西”，并且认为这些模型与现实世界并无二致，因而可以说“我们的知觉就是与现实恰好相符的一系列幻想”。同样的，Hohwy (2013)也认为“有意识的经验就像一种幻想或虚拟现实”，或许正因如此，知觉从某种意义上来说是“间接的”，大脑中的生成模型是“脆弱的自然之镜” (the fragile mirror of nature)。在这种解释中，认知是限于颅内的，似乎发生于有机体颅内的一切便可以对心理经验承担全部的责任。Hohwy 在后来的研究中以“证据边界” (亦称“感觉毯”)为限，所有认知都发生在系统证据(即感觉上皮细胞的活动)与大脑整体模型下产生的预测之间的相互作用中。在他看来，“任何与世界的紧密联系都必须在主体和世界之间的严格边界中发生”，大脑是“自证的” (self-evidencing)的，因而在某种意义上“进行推理的大脑是与世隔绝的” (Hohwy, 2016)。当然需要澄清的是，这并不意味着我们大脑中所产生的假设完全脱离世界。换句话说，尽管证据边界的内部与外部是严格划分的，只有内部才能进行推理，而边界之内的统计模型与感知内容是有联系的，比如，降雨平均值与获取实际降雨量既是有所不同的但又是有关联的。感知是一种由“果” (大脑的内部状态)溯“因” (即外部世界的隐藏诱因)的推理过程。

如果预测大脑的确以“证据边界”为界限，心智被隔离在推理纱幕之后。我们自然会问：颅内

之心究竟如何认识颅外之物？脑中的模型与现实是相符的吗？大脑作为一种物理属性的事物究竟是怎样联系于非物理属性的心智的呢？要回答这些问题，首先需要思考的是心智就其本质而言是什么？哲学家 Levine (1983)以“疼痛与 C-纤维的激发”为例，指出神经活动的解释虽然它在生理上可能是有效的，但它无助于我们理解疼痛的感觉，它们之间存在着“解释鸿沟” (the explanatory gap)。PP 内在主义版本认为心智状态在认识过程中具有通达的优先性。有研究者以一种引人注目的方式指出，预测误差最小化理论是“对简单笛卡尔怀疑论的肯定” (Hohwy, 2016)，换言之，我们接收到的感官输入并非我们通常相信的外部世界所引发，而有可能是邪恶的科学家造成的，这种怀疑论假设被预测框架所采纳。如所周知，笛卡尔采用“普遍怀疑”的方式否定了我们关于外物的素朴观念，并由此返回到“我思”这个内在精神世界，对他来说“没有什么东西比精神更容易被认识” (Descartes, 1993)，他在第二个沉思中以蜡块为例，将蜡拿到火炉旁，它的气味、颜色、形状、尺寸等物理性质都发生了较大的改变，我们仍然承认它是同一块蜡。笛卡尔觉察到，我们感官注意到的那些东西并非感知的决定因素，只能通过理智或者精神才能理解或领会这块蜡，对这块蜡认识的说明也适用于所有外在于我的东西上。这说明，我们以先前经验所形成的内在模型为基础，在脑中构成了关于外部世界的影像，心智在感知过程中起到了关键作用。或许正因如此，研究者将内在主义的预测心智称作是“笛卡尔式的” (Williams, 2018)，也就不足为奇了。

### 3.2 认识的不可超越困境

循以上思路，如果PP采纳笛卡尔怀疑主义的观点，遵循严格的内在主义立场，我们将面临以下疑难：其一，怀疑主义迫使人们认为，外部世界(包括身体)的存在与否是无关紧要的。由于对输入的感觉信息的理解只能完全依赖于证据边界之内的状态，唯有“脑中的世界模型”是真正给予我们的，进而言之，我们的身体也是证据边界之外的隐藏原因，因而，证据边界之内如何认识边界之外的对象成为一个谜。其二，PP 模型以内在神经机制来说明认知过程似乎有正当理由，而知觉具有一种超越的性质，当心智问题被暗中巧妙地转化成了神经过程，有意义的知觉过程是很难知道的。当我们“听”声音，它可能不止是一种声音感觉而是一首有旋律的乐曲；当我们“看”某物，我们虽然并不能从某个角度看到它的每一个面，但却可以“预测”到它是一张桌子，在这听或看的过程中，究竟发生了什么？时至今日，在脑成像技术逐渐变成认知科学研究的家常便饭的时代里，这种技术使得不可见的身体内部成为可见的，这无不令人叹为观止，也足以见得，人们已经习惯于认为心理的经验取决于大脑的精巧工作，其中往往隐含着一种假定，即大脑(brain)神经活动就是心智(mind)状态。但是，我们应当谨记美国心理学之父詹姆斯语重心长的告诫，“绝不要忘记被我们作为起点的自然科学的假定是暂时的、可以修正的东西” (Pylyshyn, 1984)。其实，詹姆斯在《心理学原理》中已经洞察到，伴随着意识的出现，“一种全新的质(an entirely new nature)” (James, 1983)似乎悄悄地产生了，而这并非通过观察大脑的神经元活动就可以看见或理解的。这种质相当于一种现象意识，McClelland (2017)指出，在意识科学研究的悲观主义者看来，它“提出了一个独特的‘困难’问题”，并且是“无法克服的”。此外，Gregory 作为 PP 的先驱人物也感叹道，“通过视网膜上的刺激模式，我们感知到了物体的世界，这简直是一个奇迹” (Gallagher & Zahavi, 2021)。就目前的研究而言，PP 不太可能在这个“奇迹”的研究上取得重大突破，Schlicht 和 Dolega (2021)认为“它仍然太模糊”，并且断言“该框架也不适合提供统一的意识理论”。

对以上两个疑难问题的破解，关键在于对意识的研究工作，而现象学家胡塞尔“对意识的内容、结构、及其活动的样式和成就所进行的那种极其细致的系统分析工作，是无与伦比的” (高申春, 2012)，我们或许可以从现象学中找到启示。上述的分析可以发现，PP 模型的内在主义解释是对怀疑论的某种回应，而现象学是对近代哲学认识论背景中怀疑论的根本动机的回应。胡塞尔的现象学思想与笛卡尔的“怀疑”紧密相关，二者的基本思路是类似的，即从内在论的立场出发，将内在意识当作认识论的可靠开端。不过，胡塞尔认为二元论的问题是将“我思”当作一个实体。更严重的问题是，当笛卡尔无法解释我思与外部世界(超越物)之间的联系时，把问题推给全能的上帝。与此不同的是，胡塞尔现象学的理论根本无意于

成为内在与外在、意识与实在之间的桥梁, 在他看来, 正是“认识的超越” (Husserl, 1999)问题使得我们反思认知时陷入窘境。他敏锐地洞察到内在和超越具有两种含义, 第一种是指, 主体超越“内在”切中客观的外在对象; 第二种是在悬搁之后, 超越明证给予的东西而达到不能被直接看到或把握的东西, 对此超越的探讨蕴涵着意义赋予之过程。毕竟人类心智从事着创造性的活动, 这个活动是对源初意识活动逐步远离的过程, 它使得“现象的单纯在场获得了一种确定的‘意义’, 获得了一种特殊的、观念化的内容” (Cassirer, 1953)。当然, 认识超越的两种含义并非毫无关联, 显然前者在某种程度上掺杂了后者。它们所涉及的问题可以被概括为认识超越何以可能, 二者也分别对应于前面提到的PP所面临的两个疑难, 这是认知之谜, 更是认知研究的动力。

## 4　PP 模型外在主义解释及其困境

### 4.1　PP 模型的认知外在主义解释

与内在主义不同, PP 的另一种解释倡导以行动导向的表征, 参与世界的行动是对神经信息处理的补充, 确立了心智状态的外在性, 并且该框架“对知觉的解释在性质上都是(认识论)外部主义的” (安迪·克拉克, 2020)。尽管神经加工过程是围绕预测误差最小化程序来进行组织的, 但 Hohwy 的“虚拟现实”是具有误导性的, 它暗示了知觉与现实的隔离。这种心物隔离的观点至少犯了两个错误：一是在主体与世界之间引入“表征纱幕”; 二是忽视了行动在预测过程中的作用。在 Clark 看来, 知觉并不是与身体和环境毫无关联的颅内过程。PP 本身对具身心智的核心思想并不相悖, 预测机制并非“在主体与世界之间添加令人忧虑的障碍”, 而是“提供了让结构化的世界第一时间进入视野的必要手段” (安迪·克拉克, 2020)。由此观之, 其对知觉的解释与上述的内在主义迥然相异：首先, 知觉是“直接的”, 我们所知觉到的东西本身并非一个假设(或世界模型, 或虚拟现实), 我们在日常生活中看见的是桌子或椅子之类的物体, 而非光在视网膜上的成像; 其次, 知觉是有固定组织、并朝向外界的, 知觉内容揭示了一个结构化的(即在给定先验知识的前提下)外部世界。

同时我们也会发现, 在内外部主义的争论中, 研究者以不同的方式使用了马尔科夫毯概念。与认知颅内主义的解释相反, Fabry 和 Clark 都否认马尔可夫毯可以标志心智与世界的界限。Fabry (2017)主张“由于感觉和主体状态之间的直接因果关系, 生物体正在与其环境不断相互作用”。Clark (2017)采取了类似的立场, 该毯本身“作为一个不断变化的过程来实现”, 任何以其为界的组织“都可以通过与内部和外部环境的交互来改变、重新部署或重新配置”。这种持续的相互因果关系表明, 该毯并非心物之间的静态的固定边界, 而是动态的可变边界。

根据 Clark 的看法, PP 框架体现了“延展认知” (Clark & Chalmers, 1998)的可能性, 即外在于生物体的结构和操作有时会成为主体认知不可或缺的成分。具体而言, 我们可以借助外在于主体的设备以获得信息, 在这一情形中, 旨在降低预测误差的多层级概率生成模型会选择那些涉及使用外部资源的行动。行动有助于将一些信息卸载到环境中, 手机和计算机等事实上代替大脑为我们有效减轻了认知负荷。安迪·克拉克(2020)断言, “我们是使用外部资源, 还是移动自己的感受器和效应器, 其实只是同一策略的不同具体表现罢了。”这种观点也得到 Kersten (2022)的支持, “可以使用 PP 故事中的元素作为‘认知标志’来构建延展认知的有效案例”。以蜘蛛网为例, 马尔科夫毯的外部状态被认为是引起蜘蛛感官观察的任何原因, 这意味着毯子的活动和感觉状态是由蜘蛛网构成, 蜘蛛网的主要功能之一是提醒生物注意潜在猎物的存在, 它扩展了蜘蛛的感官, 使其能够从环境中推断外部(隐藏)原因。他认为任何有助于自证的资源, 无论是生物体物理边界的内部或外部, 都可以包含在主体的马尔科夫毯范围内。

### 4.2　认知标志的模糊性困境

预测机制的外在主义观点至少在表面上克服了认知颅内主义的局限性。通过有机体的行动, 预测加工框架将大脑、身体和世界的不同元素连接起来。无可否认, Clark 等研究者的确发现了 PP 内在主义解释的缺陷, 从这个意义上来说, 他大概是一个尚好的问题发现者, 然而他将预测框架与具身心智、延展认知的观点相结合, 将身体和环境因素纳入了对人类认知本质的解释中, 问题在于：

首先, 研究者借助于行动、动态的马尔科夫毯将 PP 的认知范围扩展到了环境之中。考虑身体

和环境的因素对于认知的作用，是无可非议的。然而在对认知现象的理解中存在着“微妙的转移”，以导致了“耦合−构成谬误” (Adams & Aizawa, 2008)，也就是说，仅因为一个过程X (如，使用纸和笔)与一个认知过程(如，计算大数之和)有着相互作用，就认为X本身就是认知过程Y (或是认知过程Y的一部分)。然而，纸和笔所发挥的计算功能并非就等同于认知过程。此外，延展认知系统假设与延展认知假设之间的差别也是值得注意的，前者将脑、身体和环境共同视为一个认知系统，而后者宣称认知从大脑延展到身体和环境之中，是否可以从前者推出后者也是有待说明的。Fabry 和 Clark 是对的，脚趾是人(认知系统)的一部分，我们的认知系统从脑和中枢神经系统的边界扩展到了身体，纸笔用来计算大数之和，我们的确与外部环境发生相互作用。然而，由此认为认知加工过程也扩展到了脚趾上或外部环境中便显得唐突了，即通常来讲我们并不会说脚趾或纸笔上所发生的完全就是我们脑中所产生的认知加工过程。最终，上述的论证逻辑迫使我们去追问何为认知，这也是更为根本的问题。

对于认知的研究，我们可以从两个方面加以说明：第一，从认知系统之构成成分来说，外在主义借由动态可变的马尔科夫毯能够随意转换认知的边界以游走于大脑、身体和世界，然而无论如何，主观的认知现象被还原为与之异质的物质，如大脑或身体某部位的状态甚至是非生命的物质上，这无力反驳还原论的指责。第二，就人类认知的本质而言，他们回避了对何为认知这一问题的回答，显然的事实是，马尔科夫毯覆盖范围的变化与预测大脑认知的延展是同步的，认知的边界是变动不居的，认知可以发生于脑内或者在脑外(如前述的脚趾、笔纸等)，甚至在 Kersten 笔下的蜘蛛(网)等非人类生物体上进行。即便这些概念克服了颅内认知的困境，但是在这种延展之后的认知中，仍然无法回答何为人类认知、以及身体与环境的互动关系问题，即人类经验究竟如何产生的问题。“经验”被莫里斯·梅洛-庞蒂(2021)描述为“一个有限主体与一种它从中涌现且仍然参与其中的不透明存在的沟通”。他在胡塞尔的基础上提出了身体现象学，对身体行动的强调是其思想的重要特色，它无疑为认知研究的行动转向奠定了有力的思想基础。Gallagher 和 Zahavi (2021)明确指出“我们将看到 Merleau-Ponty 提供了现象学如何在认知科学中发挥重要作用的最佳例子之一。”

## 5 现象学对 PP 认知模型的启示

通过对 PP 认知模型内、外在主义解释的剖析，我们发现将认知看作是限于颅内的心智状态，抑或是处于身体和环境之中的，这种认知的绝对的内与外的说明似乎并不确切，要厘清关于认知边界问题，无可避免地要诉诸对认知本质的阐释，而体现在现象学认识批判中的核心内容恰好与此诉求相切合，本部分结合现象学的资源试图找到根本问题并提供可能的思考路径。

就内在主义的观点而言，一个根本的困难在于，它们承认身体和外部世界的预先客观存在，而我们所直接把握到的是证据边界之内的关于身体和世界的模型。毫无疑问，脑中的模型永远不可能通达原本更不可能代替原本，“认知不是通过精确的心理表征‘恢复’世界” (叶浩生 等, 2019)，所以内在主义的解释使得我们很难理解心智与世界之间相互联系的奇迹。边界的内部与外部之间产生了无法跨越的鸿沟，显然这是前述所提及的认识超越问题，也是一个在自然态度之中产生的问题。而在现象学的思维方式中，胡塞尔通过现象学的还原，排除了对一切外物客观存在的预设并将它们暂时悬置起来，借此回到绝对自明的原初意识体验。认识中超越物不象实在论所假定的那样是一种外在于我们而独立自存的东西，并且这种超越物也不是笛卡尔怀疑论所认为的一种幻象。如此来看，首先，这可以规避前文中第一种认识超越的困难，现象学所要追问的并不是主体的意识如何与外在对象相对应的，因为对象不在“意识”之外，并且在一个纯粹被直观之物的绝对自身被给予性意义上被给予，Husserl (1999)明确指出“认识现象本身是无疑的，它不具有超越之谜”，换句话说，要使认识的本质得以明晰，只有当它在直观中把其本来的面目给予我们。其次，对于认识超越第二种含义的理解，与其将预测机制看作是神经加工过程，不如将意识的意向性、意义与预测结合起来。胡塞尔在布伦塔诺的基础上进一步阐发了作为意识之本质的“意向性” (intentionality)特征，这个本质特征的关键价值在于，“通过胡塞尔关于‘事物’在意识的直接被给予中的‘呈现’而揭示出意识的‘意义授予’功能，并由

此引出意义的根源问题。” (高申春, 2012)举例而言，我们从某个角度看到桌子，那些未被看到的面已作为具有意义的东西被预测到，通过这种预测我们在认识中可以超出自身而把握到整体的意义。这说明我们会根据对事物的立义进行预测，在认识过程中自身被给予之物显现着，“同时以某种方式‘被立义’、‘被统摄’，而被我们称为对象之显现的东西，恰恰就处在这种对对诸感觉的赋灵(beseelend)立义的现象学特征中。” (Husserl, 2001)被称为桌子的某物并非感性材料本身，它在不同的认识过程中可以显现为某种四条腿的物体。对于同一个感觉内容，我们可以“这一次做这样的‘立义’(auffassen)，另一次做那样的‘立义’” (Husserl, 2001)。而且我们可以围绕桌子转圈，尽管它作为客体有不同的面，但它可能仍被知觉为同一个意义，即它被立义为桌子，在这个意义上，我通过身体行动认识到了世界。

相比之下，外在主义强调以行动为导向的预测，这种观点正确地指出了心智之于身体、世界的相关性，但它的错误在于就此匆忙地延展了认知，从而误解了人类认知的本质。这里主要可以从以下两个方面进行考虑：第一，该立场认为身体与环境的交互作用尤为重要，因为它们有助于将不必要的表征卸载到脑外以便进行更为经济的认知。安迪·克拉克(2020, p.306)宣称“预测加工框架无需过分依赖计算成本高昂、表征丰富的策略，而是有大量(更快、更贱、更具身的)替代方案或辅助手段”。而对于认知的说明，现象学号召我们“回到‘实事本身’”，即在直观中获得“明见性” (Husserl, 2001)。依胡塞尔之见，实事本身不是人们向来所默认的客观世界，也不是克拉克的“替代方案或辅助手段”，更不是康德意义上的物自体。回到实事“可以被视为对回归感性世界的呼吁，而感性世界是任何科学概念化和表达的前提” (Gallagher & Zahavi, 2021)。第二，该立场强调身体行动的作用，但对预测大脑的改造也未特别关注到身体本身的性质，而着重于如何打破该机制内在主义的困境。为改善颅内认知的局限，这个预测系统将不仅在有生命的身体实现，甚至还可能借由无生命的物质得以建立，而身体也被打上了一种特殊的物质化标签。而在现象学的视角中，身体获得了与生物学意义上的客观身体所不同的性格。胡塞尔对身体(Leib)与躯体(Körper)进行了区分(Reynolds, 2020)，通常被理解为我们主观经验的身体和物质对象的身体之间的差异。对于身体那难以把捉的迷一般的性质，现象学家曾言及，之所以我们难以弄明白意义何以占据分子结构或细胞组织，是因为内在于或产生于生命身体的一种意义贯穿到了整个世界。故而，极端的外在论自然也是不可取的。于是，有研究者试图以“多尺度整合”的认知系统边界来“超越内在主义和外在主义” (Ramstead et al., 2021)，而对于身体行动和认知究竟该在什么意义上被确立，在预测框架下何时改变身体行动何时调整认知，他们尚未给出清晰且具有说服力的回答。在现象学的羽翼下，身体无疑在认知和行动中起到了重要作用，如梅洛-庞蒂所言，“当我的知觉尽可能地向我提供一个既变化多样又清晰连贯的场景时、当我的各种运动意向通过自身展开而从世界那里接收到它们所期待的那些反应时，我的身体就与世界相联系了。” (Dreyfus & Taylor, 2015; 莫里斯·梅洛-庞蒂, 2021)。生成(enactive)进路主张认知是“生成的” (Gallagher, 2019; Varela et al., 1991)，认知既不是限于颅内，也不是仅仅简单地嵌在并被限制在周围的世界中，它也参与世界的生成。我们应当突破主客二分的思维框架，“重新找到我的主体性的事实和处于初生状态的客体，重新找到各种观念以及各种事物在其中诞生的原初层” (莫里斯·梅洛-庞蒂, 2021)，我们或将真正发现按其本性之所是的认知概念及其历史形成过程。

总之，预测加工作为认知科学的一种可能的新范式，的确为理解人类认知和行动提供了一些新的概念和独特的视角。如前所述，预测误差最小化框架凸显了预测的关键作用，且巧妙地结合了自上而下与自下而上的信息加工方式。而无论是该框架的内在主义或外在主义的立场，都在不同程度上将马尔科夫毯作为认知的边界，并或多或少地将认知还原为大脑的神经加工过程。在现象学家看来，在此意义上主体似乎也成为了诸客体中的一个客体，这不仅仅是遮蔽了一些重要的东西，还以一种不恰当的方式处理了它们的解释。认知心理学的研究至少可以获得如下的启示：一方面在研究方法上，探寻回到“实事本身”的现象学方案，而非奢望从脑神经的血流量图中“看见”意识经验是怎样产生的。另一方面在理论上，采用现象学的思维方式而非严格且盲目地生搬硬

套有关绝对的内、外主义的观点，认知的生成并非由主观预期或外部刺激的任何一方所决定，同时要认识到身体是生物结构和经验结构的同一，身体的行动使得有机体在经验交织的世界中开启了无限的可能性。现象学无疑为 PP 认知模型的研究提供了反思空间。

## 参考文献


安迪·克拉克. (2020). *预测算法:具身智能如何应对不确定性* (刘林澍 译). 北京:机械工业出版社.

高申春. (2012). 胡塞尔与心理学的现象学道路. *南京师大学报(人文社会科学版),* (1), 90–96.

莫里斯·梅洛-庞蒂. (2021). *知觉现象学* (杨大春，张尧均，关群德 译). 北京：商务印书馆.

叶浩生，曾红，杨文登. (2019). 生成认知：理论基础与实践走向. *心理学报, 51*(11), 1270–1280.

Adams, F., & Aizawa, K. (2008). *The bounds of cognition*. Chichester, West Sussex: Wiley-Blackwell.

Bruineberg, J., Dolega, K., Dewhurst, J., & Baltieri, M. (2021). The emperor's new Markov Blankets. *Behavioral and Brain Sciences, 45*, Article e183.

Cassirer, E. (1953). *The philosophy of symbolic forms* (Vol.1, R. Manheim, Trans.). New Haven: Yale University Press.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204.

Clark, A. (2017). Busting out: Predictive brains, embodied minds, and the puzzle of the evidentiary veil. *Noûs, 51* (4). 727–753.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, *58*(1), 7–19.

Descartes, R. (1993). *Meditations on first philosophy (*3rd, D. A. Cress, Trans.). Indianapolis: Hackett Publishing Company.

Dołega, K., & Dewhurst, J. E. (2021). Fame in the predictive brain: A deflationary approach to explaining consciousness in the prediction error minimization framework. *Synthese, 198*, 7781–7806.

Dreyfus, H., & Taylor, C. (2015). *Retrieving realism*. Cambridge, MA: Harvard University Press.

Fabry, R. (2017). Transcending the evidentiary boundary: Prediction error minimization, embodied interaction, and explanatory pluralism. *Philosophical Psychology*, *30* (4), 391–410.

Friston, K. (2010). The free-energy principle: A unified brain theory. *Nature Reviews Neuroscience, 11*(2), 127–138.

Friston, K. (2013). Life as we know it. *Journal of the Royal Society Interface, 10*, 1–12.

Frith, C. (2007). *Making up the mind: How the brain creates our mental worlds*. Oxford: Blackwell.

Gallagher, S. (2019). Precis: Enactivist interventions. *Philosophical Studies, 176*(3), 803–806.

Gallagher, S., & Zahavi, D. (2021). *The phenomenological mind* (3rd ed.). New York: Routledge.

Husserl, E. (1999). *The idea of phenomenology* (L. Hardy, Trans.). Dordrecht, Boston, London: Kluwer Academic Publishers.

Husserl, E. (2001). *Logical investigations* (vol.2, J. N. Findlay, Trans.). New York: Routledge.

Hohwy, J. (2013). *The predictive mind*. Oxford: Oxford University Press.

Hohwy, J. (2016). The self-evidencing brain. *Noûs, 50*(2), 259–285.

Hohwy, J. (2020). New directions in predictive processing. *Mind and Language, 35*(2), 209–223.

James, W. (1983). *The principles of psychology*. Cambridge, MA: Harvard University Press.

Kant, I. (1996). *Critique of pure reason* (W. S. Pluhar, Trans.). Indianapolis: Hackett Publishing Company.

Kersten, L. (2022). A new mark of the cognitive? Predictive processing and extended cognition, *Synthese, 200*(4), 1–25.

Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly, 64*, 354–361.

McClelland, T. (2017). The problem of consciousness: Easy, hard or tricky?. *Topoi, 36*(1), 17–30.

Metzinger, T. (2017). The problem of mental action- predictive control without sensory sheets. In T. Metzinger, & W. Wiese (Eds.), *Philosophy and predictive processing* (Article 19). Frankfurt am Main: MIND Group.

Nave, K. (2022). Visual experience in the predictive brain is univocal, but indeterminate. *Phenomenology and the Cognitive Sciences, 21,* 395–419.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Francisco, CA: Morgan Kaufmann Publishers.

Piekarski, M. (2021). Understanding predictive processing. A review. *Avant, XII* (1), 1–48.

Prinz, J. (2019). Ways of mindmaking. In M. Colombo, E. Irvine, & M. Stapleton (Eds.), *Andy Clark and his critics* (pp. 222–237). Oxford: Oxford University Press.

Pylyshyn, Z. (1984). *Computation and cognition: Toward a foundation for cognitive science*. Cambridge, MA: MIT Press.

Ramstead, M. J. D., Kirchhoff, M. D., Constant, A., &Friston, K. (2021). Multiscale integration: Beyond internalism and externalism, *Synthese, 198* (Suppl 1), S41–S70.

Rao, R. P. N. & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2* (1), 79–87.

Reynolds, J. (2020). Embodiment and emergence: Navigating an epistemic and metaphysical dilemma. *Journ Transc Philosophy, 1*(1), 135–159.

Schlicht, T., & Dolega, K. (2021). You can't always get what you want: Predictive processing and consciousness. *Philosophy and the Mind Science, 2*, 1–25.

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences, 17*(11), 565–573.

Seth, A. K. (2021). Preface: The brain as a prediction machine. In D. Mendonça, M. Curado, & S. S. Gouveia (Eds.), *The philosophy and science of predictive processing* (pp. xiv-xvii). London: Bloomsbury Academic.

Swanson, L. R. (2016). The predictive processing paradigm has roots in Kant. *Frontiers in Systems Neuroscience, 10* (79), 1–13.

Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. Cambridge, MA: MIT Press.

Wacongne, C., Changeux, J.P., & Dehaene, S. (2012). A neuronal model of predictive coding accounting for the mismatch negativity. *Journal of Neuroscience, 32*(11), 3665–3678.

Walsh, K. S., McGovern, D. P., Clark, A., & O'Connell, R. G. (2020). Evaluating the neurophysiological evidence for predictive processing as a model of perception. *Annals of the New York Academy of Sciences, 1464*(1), 242–268.

Williams, D. (2018). Pragmatism and the predictive mind. *Phenomenology and the Cognitive Sciences, 17* (5), 835–859.

# Predictive processing and the bounds of cognition: A phenomenological perspective

LI Xueyu, GAO Shenchun

(*Jilin University School of Philosophy and Sociology, Changchun* 130012, *China*)

**Abstract:** Predictive processing is now a prominent theoretical framework in cognitive science and philosophy of mind. This framework has two different ways of understanding the bounds of cognition, which are divided into internalism and externalism. The former claims that it is imperative to regard Markov blanket as an evidentiary boundary between the brain and its environment, and identify cognition as an internal mental activity, which is a kind of neuro-representationalism; The latter advocates the combination of this framework with the idea of embodied mind and extended cognition, and the action-oriented representation makes there is no static and fixed boundary between mind and matter. Both of them are faced with unexplainable difficulties. Internalism encounters the dilemma of transcendence of cognition, while externalism suffers from the fuzziness of the mark of cognition. Phenomenology has made a meaningful exploration on the nature of cognition, which undoubtedly provides a reflective space for the cognitive model of predictive processing.

**Keywords:** prediction processing, the bounds of cognition, internalism, externalism, phenomenology